9. CONCLUSION

The thesis titled “Graph Based Segmentation and Deep Learning for Phoneme Pattern Classification in Tamil Continuous Speech” elucidated the research work carried in developing phoneme recognition models to classify the phonemes in Tamil continuous speech. Pattern recognition had shown its prominence in various application areas. Speech to text engines, speech enabled search engines, isolated word recognizers, digit recognizers, automated customer support systems are few applications evolving in Tamil speech recognition. The higher variability in the speech characteristics of the people requires efficient machine learning models to deal with. This research work explicated the generative modelling techniques like Artificial Neural Networks, the hybrid ANFIS model and deep belief networks to solve the phoneme recognition problem. Tamil phoneme recognition models have been experimented using DBN back propagation with contrastive divergence pre-training, DBN back propagation with PSO based pretraining DBN back propagation with a proposed TPSO based pre-training have been proposed and developed. In addition, a weighted mean square error loss function has been proposed and verified by building DBN, PSO-DBN and TPSO-DBN models for Tamil phoneme recognition.

Due to the lack of Tamil speech corpus, an effort to build a Tamil speech corpus “Khazhangiyam” has been undergone. The corpus building process is supported with a graph cut based segmentation algorithm that has been proposed and developed to segment the continuous speech into phonetic units. The Kazhangiyam Tamil speech corpus has been developed with a collection of Tamil sentences that span 39 different phonetic sounds of Tamil language, which is spoken by 39 native Tamil speakers. The speech corpus is further expanded to comprise a phoneme dataset named DWTFS and used throughout the research. The Graph cut based segmentation algorithm to segment phonetic units from the Tamil continuous speech have been proposed and implemented which transforms the continuous speech into a graph and uses the statistical measures in identifying the candidate values to perform the graph cut. The graph cut based segmentation showed 13% improved precision when compared other state of art methods considered.

A pilot work was undergone with ANN and ANFIS to evaluate their efficiency to the Tamil phoneme recognition problem under consideration. The ANN phoneme recognition model was built with one hidden layer, where the training involved forward pass and backward pass. The forward pass propagated the energies of neurons in previous layers to the output layer, whereas the backward pass propagated the error observed in the output layer to the previous layers to fine tune the network parameter, thus inducing the learning process. The next recognition model built using ANFIS took the output of LDA as input and learnt itself to build the phonetic classifier. LDA was applied on the DWTFS dataset to undergo a dimensionality reduction by performing a transformation of input space to a new space and prepare the dataset suitable for ANFIS model building. During the experiments ANN was found to outrank ANFIS through is performance and generalization capability.

Inspired with the results of ANN in capturing the speech variability in a better way when compared to ANFIS, further research proposed and implemented a phoneme recognition architecture using DBN, a deep learning architecture capable enough to capture complex features. Thus, a DBN acoustic model to recognize the Tamil phonetic units of the continuous Tamil speech was built with contrastive divergence pretraining and backpropagation training phases. The efficiency of the DBN model of various depths have been studied and found to achieve an improved accuracy for the DBN acoustic models with an increase in the depth of the network. But, a trade-off of huge increase in DBN training time was observed. Eventhough, the accuracy of the DBN acoustic model improved with depth, a declination in accuracy was observed after a particular threshold in depth of DBN.

A PSO-DBN framework was proposed and used PSO pre-training to identify the optimal values for initialising the parameters of DBN. The DBN parameters were further fined tuned using back propagation based training technique that propagated the MSE observed at output layer backward through the network. This framework had shown an improvement of 3% in terms of accuracy to DBN acoustic model built for the Tamil phoneme recognition. In addition the problem of increase in training time as a multiple of depth of the DBN in contrastive divergence learning observed in previous work was significantly reduced in PSO based training.

A new improved version of PSO namely, TPSO was then proposed and took the place of PSO in PSO-DBN framework to pre-train the DBN model as an effort to solve the problem of risk of stagnation in particle movement observed in PSO. This was handled by the introduction of temperature term in TPSO that also paved way to reach the optimal solution for the DBN parameter optimization problem in fewer generation of TPSO pretraining. It also has solved the problem of particle stagnation observed in PSO based pretraining. The better optimal solution for initial DBN parameters have reflected by achieving about 90% recognition rate for the TPSO-DBN acoustic model which outperformed when compared to the earlier models built using DBN and PSO-DBN.

A loss function termed weighted mean square error was proposed to dynamically handle the influence of classes in imbalanced dataset in the model building process. The loss function has been tested and verified for the DBN, PSO-DBN and TPSO-DBN acoustic models for DWTFS phoneme dataset and found to build better models when compared to the models that were built using MSE as loss function in pre-training and training phases of DBN, PSO-DBN and TPSO-DBN. Thus, efforts that were put forth to develop efficient acoustic models for recognizing the phonetic units of Tamil continuous speech attained a level of worth.

The findings of the research are summarized below.

- Application of the dynamic algorithm, graphcut based segmentation help in properly identifying the boundaries of the speech units and thus paves path to segment the continuous speech to the required phoneme level units.
- The use of statistical measures in deciding the cut points in graphcut based segmentation algorithm speeds up the process of segmenting the continuous speech in addition to improving the accuracy of the phoneme segmentation process.
- Development of ANN acoustic model for phoneme classification in Tamil continuous speech provide better solutions when compared to ANFIS model, due to the fact that ANNs have better generalizing capability when compared to ANFIS for the problem like phoneme recognition with higher variable characteristics.
- Development of deep belief network model for building the acoustic model has the capability to self-learn the highly variable phonetic data and classify the phonetic units with accuracy comparable to the state-of-art classifiers.
- The time taken by the DBN to learn increases with the increase in the depth of the network. The first phase in training the DBN that uses contrastive divergence contributes much in the time complexity of the DBN because of its layer by layer learning procedure.
- The replacement of contrastive divergence phase in DBN training with PSO to prepare the network for back-propagation fine tuning help to drastically reduce the time complexity in building DBN based acoustic model. It also helps to reach the global optima in the solution space.
- Utilization of the proposed TPSO in preparing the DBN for back-propagation training instead of standard PSO helps in reaching the optimal solution parameters for the DBN in

fewer generations and succeeded in handling the problem of particle stagnation generally observed in PSO.

- Exploiting the proposed Weighted Mean Square Error loss function in the pre-training and training phases of DBN model building process improves the performance of the classification model.

The research contributions made through this thesis are,

- Tamil speech corpus named ‘Kazhangiyam’ has been developed.
- An improved graphcut based segmentation algorithm with statistical decision making has been designed and developed.
- A framework to automatically recognize the phonemes using deep belief networks is designed and developed.
- A framework that incorporates particle swarm optimization for pre-training the deep belief networks has been designed and developed for Tamil phoneme pattern recognition.
- A novel temperature controlled particle swarm optimization algorithm has been developed and is incorporated in the pre-training phase of the DBN framework to build Tamil phoneme recognition model.
- A novel weighted mean square error loss function has been proposed to handle imbalance dataset problem.

The thesis has uncovered few problems in building the acoustic model for Tamil phoneme pattern classification in Tamil continuous speech. As a scope for future work, the false and missing boundaries can be identified to improve results of continuous speech segmentation. In addition, this work can be implemented for other south Asian languages. It can be extended in future by building the language model to help support word/sentence level recognition of continuous Tamil speech.